

CLAIMS

B4

InsB4

Sub B5
cot

1. Apparatus for summarising data sets, the apparatus having:
an input for receiving a data set to be summarised;
sectioning means for dividing said received data set into one or more
sections according to pre-determined criteria;
ranking means operable for each said section to compare data within the
said section with one or more target data items and for calculating a ranking value
for the said section, said ranking value being dependent on the outcome of said
comparisons for the said section; and
selecting means for compiling a customised summary of the data set by
selecting one or more of said one or more sections according to their respective
ranking values.
2. Apparatus according to Claim 1, including a user input for entering target
data items.
3. Apparatus according to Claim 1 ~~or Claim 2~~ including:
means for identifying one or more key data items in each said section
according to a pre-determined stop list;
calculating means operable for each said section to calculate one or more
distribution values, each said distribution value representing a different pre-
determined measure of the distribution, in said data set, of key data items
identified in the said section; and
adjustment means for adjusting said ranking value for each said section
according to the respective said one or more distribution values.
4. Apparatus according to Claim 3, wherein said calculating means are
operable to calculate a first distribution value for each said section, said first
distribution value representing a measure of the number of sections of said data
set, other than the said section, containing key data items of the said section, said
first distribution value, as calculated for the said section, being proportional to the
sum of the values of said measure of the number of sections determined for each
key data item of the said section.

862090" E097060

a 5. Apparatus according to Claim 3 or ~~Claim 4~~, wherein said calculating means are operable to calculate a second distribution value for each said section, said second distribution value representing a measure of the separation between
 5 the first occurrence within said data set of each key data item of the said section and the respective last occurrence, said second distribution value, as calculated for the said section, being proportional to the sum of the values of said measure of separation determined for each key data item of the said section.

a 10 6. Apparatus according to ^{Claim 1} ~~any one of the preceding claims~~, wherein said selecting means are arranged to compile a summary having a pre-defined length by selecting, in order of decreasing rank, as determined by the corresponding ranking value, one or more of said one or more sections, beginning with the highest ranked section, and adding each selected section to the summary until the summary has
 15 attained said pre-defined length.

7. A method for generating a customised summary of a data set, the method including the steps of:

- i) receiving, as input, a data set to be summarised;
- 20 ii) dividing said data set into sections according to predetermined criteria;
- iii) comparing data items in each said section against one or more target data items;
- iv) calculating a ranking value for each said section in dependence upon the outcome of the respective said comparisons; and
- 25 v) compiling a customised summary of said data set by selecting one or more of said one or more sections according to their respective ranking values.

8. A method according to Claim 7, including the steps of:

- a) identifying key data items within each said section from step ii) according to a pre-determined stop list;
- 30 b) calculating, for each said section, one or more distribution values each representing a pre-determined measure of the distribution of the key data items of the said section in said data set; and

c) adjusting said ranking value from step iv) for each said section in dependence upon the respective said one or more distribution values.

9. A method according to Claim 3, wherein, at step b), said one or more pre-determined measures of distribution include a measure of the number of sections of said data set, other than the said section, containing key data items of the said section and wherein the corresponding distribution value, as calculated for the said section, is proportional to the sum of the values of said measure of the number of sections determined for each key data item of the said section.

10

10. A method according to Claim 8 or Claim 9, wherein, at step b), said one or more pre-determined measures of distribution include a measure of the separation between the first occurrence within said data set of each key data item of the said section and the respective last occurrence and wherein the corresponding distribution value, as calculated for the said section, is proportional to the sum of the values of said measure of separation determined for each key data item of the said section.

15

Subs
Caled

a

add Be

03077603.060298